INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 20, Issue: 5, Month: October, Year: 2025
Article Number: 6911, https://doi.org/10.15837/ijccc.2025.5.6911

Enhanced Market Forecasting for Autonomous Vehicles Using a Hidden Mixture Gaussian Markov Model

Xuemei Chen, Quanfeng Qiu, Zuanxu Chen

Xuemei Chen
School of Intelligence Technology
Geely University of China, China
123 Chengjian Avenue, Eastern New District, Chengdu, Sichuan Province, China
plumblossom2024@aliyun.com

Quanfeng Qiu*
School of Intelligence Technology
Geely University of China, China
123 Chengjian Avenue, Eastern New District, Chengdu, Sichuan Province, China
*Corresponding author: ouhouo@aliyun.com

Zuanxu Chen
Research Institute of Finance and Trade
Sichuan Academy of Social Sciences, China
arrowczx@163.com

Abstract

In the context of the rapid development of autonomous vehicle technology and its increasing influence on the global automotive industry, the ability to accurately forecast market trends has become crucial for strategic decision-making. This study introduces the Hidden Mixture Gaussian Markov Model (HMGMM), a novel probabilistic framework designed to enhance the precision of market trend predictions in the autonomous vehicle sector. By addressing the limitations of traditional Hidden Markov Models (HMMs), which struggle with high-dimensional continuous data and dynamic market fluctuations, the HMGMM integrates Gaussian distributions to better capture the complexities of market dynamics. Utilizing a sliding time window mechanism and an improved algorithm for parameter dynamic updates, the HMGMM significantly improves response speed to market changes. The research employs experimental analysis on real-world datasets to validate the model's effectiveness, demonstrating superior predictive performance with an accuracy of 0.892, recall of 0.901, and reduced RMSE of 0.144. These results highlight the potential of HMGMM as a reliable tool for market trend prediction, emphasizing the need for both the automotive industry and market analysts to adopt advanced probabilistic models to anticipate future market shifts and capitalize on the opportunities presented by autonomous vehicle technology.

Keywords: Autonomous driving car; Market forecasting; Hidden Markov Model; Accuracy; RMSE

1 Introduction

The rapid development of the Autonomous Vehicles (AVs) market, propelled by a multitude of dynamic factors, presents significant challenges to predictive models due to its inherent complexity and uncertainty. Traditional forecasting methods, particularly the Hidden Markov Model (HMM), suffer from substantial limitations when dealing with high-dimensional, continuous, and nonlinear market data. HMMs, which assume discrete hidden states and static probability distributions of observed values, struggle to capture the continuous dynamic features embedded in market data. Existing studies have identified several key issues with HMMs in the context of the AVs market. The modeling of high-dimensional continuous data is problematic, as market indicators are essentially continuous time series signals. HMMs must process such data through discretization or vectorization, leading to information loss and the so-called "dimensional disaster" [1]. Moreover, the inadequate adaptability of HMMs to dynamic fluctuations is a critical concern, given that the AVs market is highly susceptible to sudden policy adjustments and technological breakthroughs [2]. Additionally, market data often exhibit multi-modal distributions, which the single Gaussian observation hypothesis of HMMs cannot effectively represent. Although recent research has attempted to enhance prediction performance through deep learning or integrated approaches, significant shortcomings remain. While traditional Gaussian mixture models can capture multi-modal distributions, they lack the capability to dynamically model time series data [3, 4]. Current research has yet to effectively integrate time dependence and probability distribution flexibility, making it difficult to achieve a balance between accuracy and robustness in market predictions for the AVs industry [5]. To address these challenges, we propose a Hidden Mixture Gaussian Markov Model (HMGMM). This innovative model can directly model continuous market data by replacing the observation probability with a multi-Gaussian mixture distribution, thereby avoiding discrete information loss and accurately capturing multi-modal features. The introduction of a sliding time window mechanism allows for the segmentation of long time series data, while an improved algorithm enables dynamic parameter updates, significantly enhancing the model's responsiveness to market changes. This research not only fills the technical gap between continuous dynamic modeling and multi-modal characterization but also offers a new paradigm that combines theoretical rigor with engineering practicality for high-precision predictions in more complex market environments.

The proposed model is evaluated through a structured approach that includes four main components. First, a comprehensive review of the literature on predictive models, probability theory, and Big Data Analytics (BDA) is conducted, critically assessing the achievements and limitations of prior research. Second, the proposed algorithm is constructed and analyzed in detail, with an emphasis on the introduction of an improved method that enhances the model's capabilities. Third, the performance of the model is rigorously verified through comparative experiments, demonstrating its effectiveness and superiority over existing methods. Finally, the experimental results are summarized, highlighting the study's limitations and suggesting directions for future research. The meanings of all mathematical symbols in the Paper are summarized in Table 1.

2 Literature review

2.1 Market Forecasting Models: models, challenges and limitations

The development of autonomous vehicle (AVs) represents a significant technological advancement in the automotive industry, promising to transform transportation systems globally [6, 7]. The market for AVs is a rapidly evolving and highly dynamic sector within the transportation industry, up to 13,632.4 billion dollars in 2030 [8]. As the AVs market continues to evolve, the ability to anticipate changes and adapt to emerging trends becomes increasingly vital for maintaining competitive advantage [9] [10, 11]. Market forecasting for AVs is a critical area of research, given the potential transformative impact of this technology on transportation systems, urban planning, and the environment, increasing road capacity and mitigate traffic congestion [12, 13]. Effective market forecasting enables the authority and companies to make informed strategic decisions, allocate resources efficiently, and mitigate risks associated with market volatility [14]. However, Market forecasting for AVs is a complex task due to the interplay of data, technological, regulatory, and consumer-driven factors.

Market forecasting for AVs is very complicated, influenced by regulatory and policy changes [16, 48], consumer acceptance and trust, consumer behavior and preferences, technological and infrastructure requirements, safety and ethical concerns, economic and environmental impacts. Rapid advancements

Table 1: Summary of meanings of mathematical symbols

Mathematical symbol	Implication
A	A random sequence in a probability space
S	The state
t	Time
a_{ij}	The transition probability from the current hidden state i to the next hidden one j
A	Transition moment probability matrix
S	Homogeneous Markov chain
Q	The set of all possible states
V	The set of all possible observations
b_{ij}	A transition possibility representing the likelihood from states i at t to j at $t+1$
A	The transition probability matrix
B	The observation probability matrix
$b_i(v_t)$	The possibility of generating the observation value v_t at time t in state i
π	The initial state probability distribution
π_i	The possibility at the initial moment $t = 0$ in the initial state i
HMM	Hidden Markov Model
S_t	The hidden state at time t
$O_{1:t}$	Observed sequence from time 1 to time t
$P(O_{1:t})$	The probability of observing a sequence
$P(O_{1:t}\|S_t)$	The possibility of the observing sequence from $O_{1:t}$ when the hidden state is S_t at time t
$S_{1:t}$	The hidden state sequence
$P(O_{1:t}, S_{1:t})$	The probability that the observation sequence and the hidden state sequence appear at the same time under the model parameters
$L(\theta)$	The log-likelihood function of the probability
θ^*	The result of solving the $L(\theta)$ function
$\theta^{(k)}$	The estimated parameter value obtained from the last iteration
$\delta_t(i)$	The maximum possibility of the path with status i at time t

in AI [10, 17], connectivity, and sensor capabilities are crucial for the growth of the AVs market. The market for AVs is heavily influenced by regulatory and policy changes [13, 18]. Forecasting models must account for these external factors, which can be challenging due to their unpredictability. [19] discusses the necessity of introducing autonomous trucks in logistics, highlighting the regulatory challenges. [20] examines the life cycle greenhouse gas emissions of transitioning to an AVs fleet, emphasizing the impact of policy changes. The widespread adoption of AVs depends on consumer acceptance and trust [21]. Forecasting models must consider the impact of consumer attitudes, which can be influenced by factors such as safety concerns and public perception [22, 23]. [23] explores the psychological and socio-demographic influences on AVs adoption in Malaysia, highlighting the importance of consumer trust. Similarly, [24] examines the factors influencing the willingness to use public AVs. However, consumer attitudes towards AVs are still evolving, and there is significant uncertainty regarding their willingness to adopt this technology [25]. The successful deployment of AVs requires significant technological and infrastructure advancements, including improvements in communication systems, sensor technology, and road infrastructure. [26, 27] discuss the challenges and opportunities of single-photon LiDAR systems for AVs, highlighting the technological requirements. Additionally,

[28] proposes a dynamic queueing model for shared AVs, emphasizing the need for infrastructure improvements. Safety and ethical concerns are major factors influencing consumer acceptance and regulatory approval of AVs. [29] discusses the use of GAN-enhanced predictive frame synthesis for AVs, emphasizing the importance of addressing safety concerns. Similarly, [30] constructs a robustness benchmark for motion forecasting, highlighting the ethical considerations in model predictions. The economic and environmental impacts of AVs are still being studied. These impacts, including changes in transportation costs, energy consumption, and pollution levels, can influence market dynamics and must be considered in forecasting models. [31] analyzes the reduction in urban air pollution due to the expansion of eco-friendly vehicles, highlighting the environmental impacts of AVs. [20] examines the life cycle greenhouse gas emissions of transitioning to an AVs fleet, emphasizing the economic and environmental considerations.

Market forecasting is complex not only because of the above issues, but also because of models' inherent uncertainty, limited historical data, dynamic market conditions, reliance on assumptions, models' own complexity. Market forecasting models have been widely researched, in spite of the challenges and limitations.

Autonomous vehicles are a relatively new technology, and comprehensive historical data on their market penetration is scarce [31, 32]. [14] highlights the challenges of establishing a reliable quantitative research approach for measuring AVs penetration due to data limitations. Similarly, [48] emphasize the need for thorough analysis to forecast the penetration of electric vehicles (EVs), which can be extrapolated to AVs. AVs-related data (e.g., sensor outputs, traffic patterns, and user behavior) is highly heterogeneous, making it difficult to integrate and analyze. The market for AVs is rapidly evolving, with continuous technological advancements and changing consumer preferences. Forecasting models must adapt to these dynamic conditions, which can be challenging due to the need for real-time data updates. [19**?**] explores the determinants of personal concern about AVs, highlighting the evolving nature of consumer attitudes. [34] discusses the growing trend of 4D scene perception and prediction, emphasizing the need for continuous updates in forecasting models.

The future market penetration of AVs is inherently uncertain due to unpredictable technological breakthroughs, regulatory changes, and consumer acceptance. Even the most sophisticated forecasting models cannot account for all possible future scenarios. The review by [35] discusses the challenges of motion prediction in autonomous driving, highlighting the uncertainties in forecasting future scenarios. Similarly, [36, 37] proposes a driving world model that aims to address these uncertainties but acknowledges the inherent limitations. [38] predicts vehicle ownership growth using the Gompertz model, highlighting the limitations in forecasting market penetration and emphasizing the need for accurate forecasting models.

Forecasting models often rely on assumptions about future market conditions, technological progress, and consumer behavior. If these assumptions are inaccurate, the reliability of the forecasts diminishes. [39] presents a method to predict EV market penetration and its impact on energy saving and CO_2 mitigation, but acknowledges the limitations due to assumptions about future technological advancements. [36] categorizes techniques for future prediction and behavior planning, highlighting the reliance on assumptions in these models.

Advanced forecasting models, such as those based on neural networks [40, 41], Markov chains or deep learning [42, 43, 44], can be highly complex in model interpretation and validation. [45] introduces a large-scale interactive motion dataset for developing joint prediction models, but acknowledges the challenges in validating these complex models. Real-time forecasting demands low-latency processing of high-dimensional data (e.g., traffic simulations, sensor fusion) [46], which strains traditional models and requires significant computational resources. [42] discusses the computational challenges of integrating deep learning models with AVs forecasting, particularly in handling large-scale data. [47] highlights the limitations of deploying complex forecasting models on edge devices, which are often used in AVs systems. [48] introduces a novel LiDAR perception task for occupancy completion and forecasting, highlighting the rapid technological advancements in the field. Additionally, [49] discusses the challenges of non-stationary spatio-temporal modeling, emphasizing the impact of technological advancements on forecasting.

2.2 HMM forecasting models

Hidden Markov Models (HMMs) have emerged as a cornerstone in sequential data analysis, leveraging their ability to model hidden states underlying observable sequences. HMMs applications have expanded across diverse domains, including healthcare, predictive maintenance, market forecasting, and big data analytics. However, inherent challenges and limitations persist, driving researchers to innovate modified HMM architectures and hybrid methodologies.

In healthcare, HMMs have significantly enhanced medical imaging and diagnostics [4]. For instance, a novel loss function integrating HMMs improved 3D medical image segmentation by penalizing implausible anatomical predictions [50]. Similarly, combining Hidden Markov Random Fields (HMRF) with the Whale Optimization Algorithm (WOA) optimized brain MRI segmentation accuracy, aiding precise treatment planning [51]. Beyond imaging, HMMs like NeuroPeptide-HMMer (NP-HMMer) advanced neuropeptide discovery in proteomics, particularly in understudied invertebrates [52].

Predictive maintenance systems benefit from HMMs' temporal modeling capabilities. A multi-channel fusion method combining HMMs with Bayesian theory enabled real-time remaining useful life (RUL) predictions for industrial tools, facilitating proactive maintenance strategies [53, 54, 55, 56]. In machine condition monitoring, Hidden Semi-Markov Models effectively recognized operational states [57, 58], while HMM-based frameworks optimized energy consumption in mobile fog computing through dynamic computation offloading [47, 59].

Market forecasting has seen innovative HMM applications, such as nonparametric HMMs that reduce modeling bias in stock price predictions by incorporating flexible emission models [60, 61]. Bitcoin price forecasting and multivariate Markov-switching models for crude oil markets further demonstrated HMMs' adaptability to volatile financial data [62, 63, 64]. Fraud detection in electronic banking also leveraged HMMs to identify suspicious transactions in imbalanced datasets [65].

Despite their versatility, traditional HMMs face accuracy limitations in tasks like heart sound segmentation and disease progression modeling. Standard HMMs struggled with noisy labels in named entity recognition, prompting the development of Conditional HMMs (CHMMs) for unsupervised label correction [54, 66]. Scalability remains another critical issue, as large state spaces strain computational efficiency. Researchers addressed this via optimized regularization techniques and scalable architectures [67]. Another challenge is computational complexity which hinders HMM integration with big data analytics. Bayesian models scaling quadratically with data size demand specialized hardware or parallel processing [68, 69]. Additionally, manual expertise dependency in risk management and parameter tuning limits automation potential [70].

Modified HMM variants have emerged to tackle these challenges. Duration HMMs (DHMMs) improved heart sound segmentation accuracy without relying on electrocardiogram data [66]. Tailored HMMs (THMMs) and Evolving Connectionist Systems (ECoS) were designed for niche applications like cellular map matching and tool wear monitoring [58, 71]. Hybrid frameworks integrating HMMs with machine learning algorithms also gained traction. For instance, combining Pearson correlation, exponential filters, and HMMs enhanced gaze-controlled object selection accuracy [72]. In big data, graph-based HMM optimization algorithms improved time-series processing efficiency [73, 74].

The integration of HMMs with other techniques such as Bayesian networks, big data analysis [75, 76], and machine learning algorithms has shown promising results in addressing complex problems. One of the key applications of HMMs and big data analysis is in fault detection and prediction of complex engineered systems [55]. Another significant application of HMMs and big data analysis is in the optimization of big data processing using graph-based approaches [73]. One of the challenges is the computational intensity of big Bayesian models that scale quadratically with the number of observations [69, 77]. This computational complexity can limit the scalability of big data analysis tasks and may require specialized hardware or parallel processing techniques to handle large datasets efficiently. Another limitation of HMMs and big data analysis is the reliance on manual labor and professional expertise for risk management tasks [70]. Thai H D et al. proposed a medical data processing method in BDA [56]. Multiple data processing architectures were used to capture key data such as pathology and distance, and analyze the captured data to build an application that provides recommended solutions, while big data analysis can provide valuable insights into safety risks and production activities

This paper explores the potential of hybrid models, especially the implicit hybrid Gaussian Markov model. By introducing Gaussian mixture model instead of the traditional single Gaussian distribution hypothesis, the multi-modal characteristics of market data can be better captured and continuous time series data can be effectively processed.

3 Research methodology

3.1 Hidden Markov Model (HMM)

HMM is typically composed of two random processes, which are based on a Markov chain observation sequence consisting of a series of potential hidden states. A Markov chain is characterized by its memorylessness, where the probability of future states is entirely determined by the current state, independent of previous event sequences. HMM includes observable and unobservable states. The latent states within an HMM do not have a direct one-to-one correspondence with the observed variables. Consequently, only considering the order of observation results cannot determine the potential hidden states that generate each observation result [78]. HMM lays the foundation for the proposed Hidden Mixture Gaussian Markov Model (HMGMM), which aims to address the limitations of traditional HMM in processing continuous data and improve the accuracy of predicting AVs market trends.

$X_0, X_1, \ldots$ represent a random sequence in a probability space, with values from countable or finite sets. The random process is a Markov chain when it satisfies equation 1.

$$P\{X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} = P\{X_{n+1} = i_{n+1} | X_n = i_n\} \tag{1}$$

$i_0, i_1, \ldots, i_n$ represents the state. Equation equation 1 is called non-aftereffect, which is a fundamental characteristic of Markov chains. This feature indicates that the state at time n is related to the state at time $n-1$, which is independent of the state before time $n-1$. The conditional probability is expressed as equation 2.

$$P_{ij}(n) = P\{X_{n+1} = j | X_n = i\} \tag{2}$$

In equation 2, $i, j \in I$. $P_{ij}(n)$ is the transition probability from the current hidden state i to the next hidden one j. The probability of the next hidden state is obtained after N different hidden states, thereby obtaining an $N \times N$ transition moment probability matrix. This matrix can describe the transition trajectory of implicit states in HMM, and each value in the matrix represents the transition probability. For any m, n, if $P\{X_n = i\} > 0$ and $P\{X_m = i\} > 0$, then equation 3 holds.

$$P\{X_{n+1} = j | X_n = i\} = P\{X_{m+1} = j | X_m = i\} \tag{3}$$

According to equation 3, $\{X_n : n \geq 0\}$ represents a homogeneous Markov chain. HMM extends the concept of Markov chains by introducing unobservable states. Therefore, it is called "hidden". These models generate sequences of hidden states following the probabilistic transitions inherent in Markov chains. Subsequently, from these hidden state sequences, a corresponding set of observable sequences is randomly produced. The state sequences delineate the interactions and transitions among the states, while the observation sequence illustrates the probabilistic link between the hidden states and the manifest observations, as governed by the observation probability distribution. This framework allows the HMM to model complex systems where the intrinsic mechanisms are not directly observable but can be inferred from the observable outcomes [79]. Assuming Z is the set of all possible states, while G is the set of all possible observations, thus $Z = \{z_1, z_2, \ldots, z_N\}$ and $G = \{g_1, g_2, \ldots, g_M\}$. A is the transition probability matrix:

$$A = [a_{ij}]_{N \times N} \tag{4}$$

In equation 4, $a_{ij} = P(s_{t+1} = z_j | s_t = z_i)$, $i = 1, 2, \ldots, N$, and $j = 1, 2, \ldots, N$. a_{ij} is a transition possibility representing the likelihood from states z_i at t to z_j at $t+1$. The matrix B in equation 5 is referred to as the observation probability matrix.

$$B = [b_j(k)]_{N \times M} \tag{5}$$

In equation 5, $b_j(k) = P(o_t = g_k|s_t = z_j)$, and $k = 1, 2, \ldots, M$. $b_j(k)$ means the possibility of generating the observation value g_k at time t in state z_j. The initial state probability distribution is represented by equation 6.

$$\pi = (\pi_i) \tag{6}$$

In equation 6, $\pi_i = P(s_1 = z_i)$. π means the initial state probability distribution. π_i refers to the possibility at the initial moment $t = 1$ in the initial state. If HMM is set to λ, HMM is represented as equation 7.

$$\lambda = (A, B, \pi) \tag{7}$$

In equation 7, these three elements in HMM are the state transiting matrix, observing probability matrix, and initial probability distribution. Markov chains and hidden states are determined by the initial probability distribution and state transition matrix, and the generation of observing sequences is determined by the observation matrix.

HMM has three fundamental challenges: probability computation, learning, and decoding. Probability calculation requires determining the conditional probability of the observed sequence given specific model and sequence conditions. The learning problem involves estimating model parameters from an observation sequence to maximize the likelihood of the observed data. In addition, the decoding task is to identify the most likely hidden state sequences that may generate a given observation sequence in the context of HMM. In terms of probability computation, the objective is to calculate the likelihood of an observation sequence occurring under a particular HMM with its parameters. This process can be approached using several algorithms, including direct computation, the forward algorithm, and the backward algorithm. The direct calculation method first determines the probability of a specific hidden state sequence, and then extends it to the corresponding observation sequence. This method provides a simple but computationally intensive approach to evaluate the observational likelihood of a given model parameter. The forward algorithm is represented by equation 8.

$$a_t(i) = P(o_1, o_2, \ldots, o_t, i_t = q_i|\lambda) \tag{8}$$

In equation 8, q_i represents the hidden state. $o_1, o_2, \ldots, o_t$ is an observed sequence from time 1 to time t. $a_t(i)$ refers to the probability of observing a sequence. There are N hidden probabilities at time T. Each hidden state corresponds to a forward probability. The final probability can be obtained by adding up each forward probability. The backward algorithm is represented by equation 9.

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \ldots, o_T|i_t = q_i, \lambda) \tag{9}$$

In equation 9, $\beta_t(i)$ represents the possibility of the observing sequence from $t + 1$ to T when the hidden state is q_i at time t. At T, $\beta_t(i) = 1$, $i = 1, 2, \ldots, N$. HMM learning is a parameter optimization problem that identifies the parameter that maximizes the possibility given a given observing sequence. Supervised learning methods are used in situations where there are hidden state sequences with complete annotations, and their training data requires high manual annotation requirements. HMM has hidden state sequences that cannot be directly observed, so unsupervised learning methods are generally used. EM is an algorithm that optimizes and iterates parameters based on given prior parameters to find local optimal results. Firstly, the HH function is solved using equation 10.

$$Q(\lambda, \overline{\lambda}) = \sum_I \log P(O, I|\lambda)P(O, I|\overline{\lambda}) \tag{10}$$

In equation 10, I represents the hidden state sequence. $P(O, I|\overline{\lambda})$ represents the probability that the observation sequence and the hidden state sequence appear simultaneously under the model parameter $\overline{\lambda}$. $\log P(O, I|\lambda)$ represents the log-likelihood function of the probability. $Q(\lambda, \overline{\lambda})$ represents the result of solving the HH function. $\overline{\lambda}$ represents the estimated parameter value obtained from the last iteration. Then, the Q function is maximized and the probability corresponding to the parameters is estimated. The decoding problem is to estimate the maximum possible hidden state sequence corresponding to a specific observation sequence, given the model parameters. The decoding problem usually uses approximation algorithms and Viterbi algorithms. The Viterbi algorithm solves the path with the highest probability based on dynamic programming. The maximum possibility of the path with status i at time t is represented by equation 11.

$$\delta_t(i) = \max_{i_1, i_2, \ldots, i_{t-1}} P(i_t = i, i_{t-1}, \ldots, i_1, o_t, \ldots, o_1|\lambda) \tag{11}$$

In equation 11, $i = 1, 2, \ldots, N$. $\delta_t(i)$ represents the maximum possibility of the path with status i at time t. Model parameters and observation sequences are inputted, and recursion is performed based on the initial state parameters. When the maximum possibility condition is reached, the recursion terminates. Finally, the optimal path is backtracked to find the optimal path.

3.2 Gaussian Mixture Model

The Gaussian Mixture Model (GMM) is a sophisticated probabilistic model that posits data points are generated from a finite mixture of Gaussian distributions with unknown parameters [44]. These distributions are convex combinations of various distributions. This model is particularly adept at capturing the underlying structure of complex multi-modal data distributions, which traditional single Gaussian models may not be able to meet. The Probability Density Function (PDF), or mixture density, is typically a weighted sum of distribution PDFs. Strictly speaking, the sum of non-negative weights is 1. Gaussian Mixture Model (GMM) is a concrete example of this mixture distribution, which assumes that data points come from a mixture of a finite but unspecified number of Gaussian distributions. The PDF of a GMM is articulated as a linear superposition of these constituent Gaussian distributions, encapsulating the collective behavior of the mixture components:

$$\sum_{k=1}^{N} \pi_k = 1 \quad (\pi_k \geq 0), \quad p(x) = \sum_{k=1}^{N} \pi_k \mathcal{N}(x; \mu_k, \Sigma_k) \tag{12}$$

In equation 12, π_k denotes the mixing coefficient for the k-th Gaussian component. $\mathcal{N}(x; \mu_k, \Sigma_k)$ is the density function of Gaussian distributions with mean $\mu_k \in \mathbb{R}^D$ and covariance matrix $\Sigma_k \in \mathbb{R}^{D \times D}$. $x \in \mathbb{R}^D$, and $N, D \in \mathbb{N}^+$.

3.3 The proposed autonomous vehicle market forecasting method based on probabilistic big data analysis

This section details the probabilistic big data model for forecasting the AVs market. Firstly, the limitations of HMM in handling continuous data are analyzed, and a Hidden Mixture Gaussian Markov Model (HMGMM) is introduced to enhance predictive ability, which involves a dynamic training method. Dynamic training has been applied in natural language processing [80, 81] and reinforcement learning environments [82], etc. Dynamic training refers to a flexible and adaptive approach to model training that allows for real-time adjustments based on the evolving characteristics of the data and the performance of the model [83]. Dynamic training continuously evaluates and optimizes training parameters, learning rates, and data inputs [85]. The core principle of dynamic training lies in its ability to optimize the training process by integrating feedback mechanisms that inform adjustments, thereby creating a more responsive and efficient learning environment. This adaptability enables models to effectively capture new information and changing conditions, ultimately enhancing their performance and accuracy, leading to more effective and precise outcomes [84].

The HMGMM integrates Gaussian models to address the shortcomings of HMM, handling continuous market data [85]. Based on the Expectation-Maximization algorithm, the model parameters are optimized. Dynamic training pools are implemented to address the long-time spans and fluctuations in market data, as a single training set may not effectively adapt to sudden changes and trend turns in the market. The model utilizes a dynamic training pool to update the training set with new market data. This process enhances prediction accuracy and response speed, ensuring the model remains adaptable and predictive in rapidly changing market conditions. The continuous updating of the training set enables the model to maintain efficiency in adapting to new market data. This results in improved predictive ability and response speed in dynamic market environments. Comparative experiments showed that HMGMM outperformed HMM and other models in accuracy and recall, providing a robust tool for market trend prediction in the AVs industry.

Market data are a typical time series data. The time attribute plays a crucial role in price changes and price state transitions. HMM cannot represent high latitude continuous data in observation, and vectorization of continuous data can lead to loss of data information. To address this issue, the study proposes HMGMM to extract and predict continuous features. Fig. 1

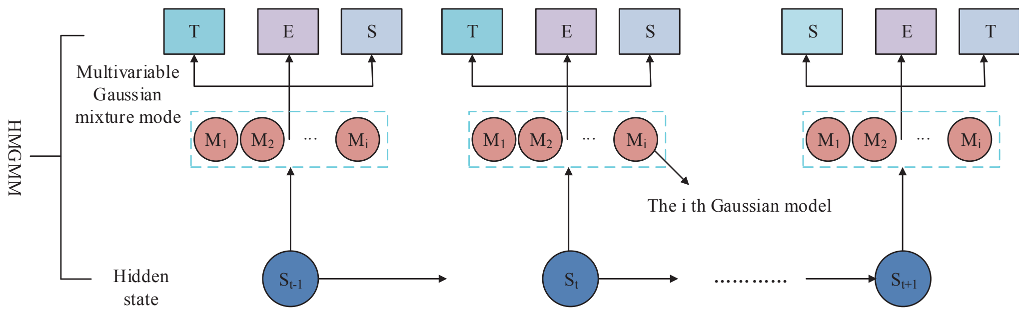


Figure 1: Structure of a Gaussian Mixture Model (GMM)

The study incorporates adjacent states into the market volatility prediction model. The state S_{t+1} at $t+1$ is influenced by the adjacent state S_t, and a Gaussian mixture model is utilized to derive the observation information O_t of the hidden state S_t at $t+1$. HMGMM is represented as a quintuple (S, O, A, B, π). $S = \{1, 2, \ldots, n\}$ represents the set of state probabilities. HMGMM utilizes the EM algorithm to search for probability maximization parameters. In general, EM is used in a single observing sequence. However, these market time series data are relatively long, and a longer time span can affect the prediction performance. Modeling as a single observation sequence will lead to lower recognition rates. The study proposes a dynamic training pool to re-estimate parameters, as displayed in Fig. 2.

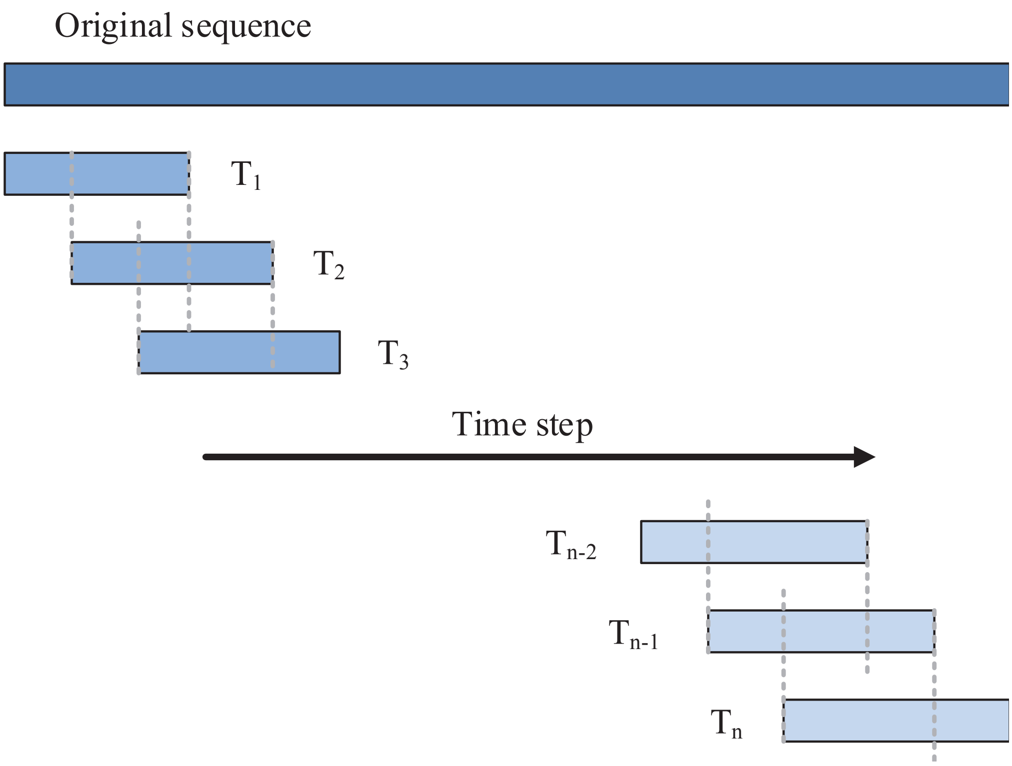


Figure 2: Dynamic training cell

Fig. 2 is a schematic diagram of the dynamic training pool, where the raw data is divided into multiple segments. The data segments are divided into appropriate lengths and time intervals to achieve the optimal window size. The complete HMGMM method operation process is shown in Fig. 3.

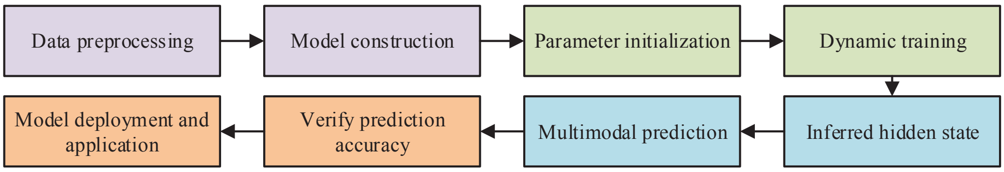


Figure 3: Operation flow of HMGMM method

4 Results and discussion

The performance analysis of the AVs market prediction model based on probability theory BDA is divided into two subsections. Firstly, the improvement effectiveness was verified. Then, it was compared with other models to verify its superiority.

4.1 Effectiveness analysis of market forecasting models

The study selected AVs data from multiple automotive websites, with data sources spanning from January 1, 2022 to January 1, 2023. After preprocessing, the data is divided into training set and test set according to 7:3. The test set serves both testing and validation purposes. The laboratory environment is set as shown in Table 2.

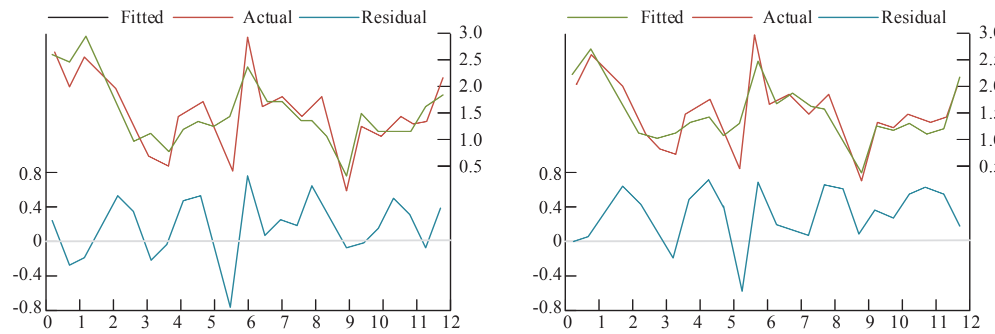
Table 2: Laboratory environment setup

Hardware and software configuration	**Version model**
CPU	Intel(R) Core i7-7700@3.6GHz
GPU	GTX 1060
Operating system	Ubuntu 18.04 LTS
RAM	32G
Display memory	6G
CUDA	9.1

Table 2 shows the laboratory environment settings. Indicators such as accuracy, recall, F1 Score, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) were selected to measure the predictive performance. MAE measures the average absolute deviation between predicted and true values, which can reflect its prediction error. RMSE measures the difference between predicted and true values, which is obtained by squared error and then squared to evaluate the prediction accuracy. MAPE evaluates the percentage of prediction error in a model, helping to understand the relative magnitude of error between predicted and actual values. HMM and HMGMM were trained on the training set. Fig. 4 shows the training results.

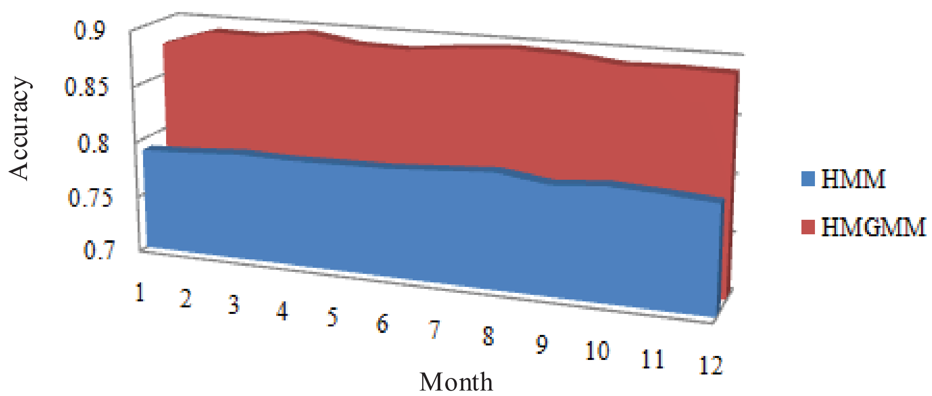
Fig. 4 (a) presents the fitting effect of HMM, and Fig. 4 (b) shows the fitting effect of HMGMM. The red line refers to the actual value, the green one refers to the fitted value, and the blue one refers to the model residual value. The overall fitting effect of these two models on the data was good, and the overall data trend was relatively synchronous. HMGMM demonstrated a better fitting effect at turning points and stronger explanatory power at poles compared to HMM. HMM and HMGMM were tested on the test set. Fig. 5 shows the accuracy and recall.

Fig. 5 (a) shows the accuracy comparison of the two models. From the results, the prediction performance of HMM was medium, and the accuracy was roughly maintained at 0.8, while HMGMM showed a high prediction accuracy, and the accuracy was stable at 0.9. This shows that HMGMM can more accurately fit market trends, reduce forecast errors, and is especially adaptable under complex

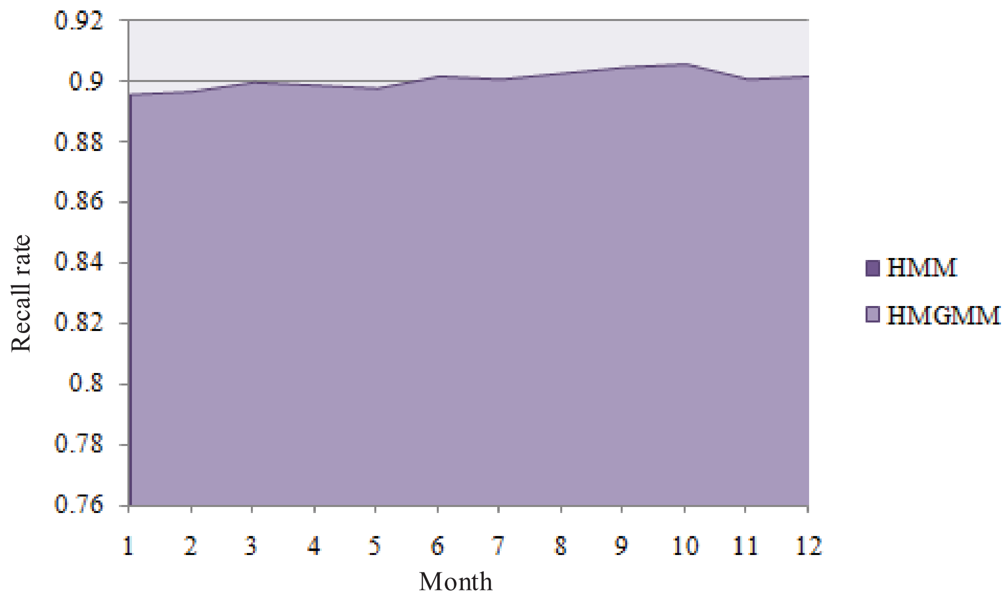


(a) HMM Model fitting effect drawing (b) HMGMM Model fitting effect drawing

Figure 4: Model fitting effect



(a) Comparison of accuracy of different models



(b) Comparison of recall rates of different models

Figure 5: Comparison of accuracy and recall rate between HMM model and HMGMM model

market conditions. Fig. 5 (b) compares the recall between these two models. The recall of HMGMM was significantly higher than HMM, indicating that HMGMM was better at identifying real changes and trends in the market, and could capture more market fluctuations and important turning points. This feature enables HMGMM to not only more accurately reflect the current market state, but also effectively predict potential future market changes and provide more valuable predictive information. HMM and HMGMM are tested on the test set as displayed in Table 3.

Table 3: Experimental comparison between HMM model and HMGMM model

Index	HMM	HMGMM
Accuracy	0.796	0.892
Recall rate	0.814	0.901
F1 score	0.756	0.897
MAE	0.158	0.122
RMSE	0.169	0.144

Table 3 shows the comparison results between HMM and HMGMM experiments. HMGMM performed better than HMM in various evaluation indicators. Compared to HMM, HMGMM had better accuracy, recall, F1 score, MAE, and RMSE. Firstly, the accuracy of HMGMM was 0.892, which was higher than HMM's 0.796. In terms of recall rate, HMGMM was 0.901, which was significantly better than HMM's 0.814. In terms of F1 score, HMGMM was 0.897, which was also significantly higher than HMM's 0.756. In addition, the MAE and RMSE of HMGMM were 0.122 and 0.144, both lower than the corresponding values of HMM. This indicates that HMGMM captures the changing trends of real data more accurately in market forecasting, with smaller relative errors and more accurate and reliable predictive performance of this model. Therefore, based on the performance of the above indicators, HMGMM performs better than HMM in market forecasting models, with stronger predictive ability and higher accuracy, making it more suitable for market trend prediction.

4.2 Performance analysis of market forecasting models

After verifying the effectiveness of the model improvement, the model performance was further tested. This study compares the Autoregressive Integrated Moving Average model (ARIMA), Long Short-Term Memory (LSTM), Prophet Forecasting Mode (Prophet), and Random Forest (RF) models that can be used for market forecasting with the proposed HMGMM. ARIMA is a classic time series analysis method that can capture the data trends and periodicity. RF is an ensemble learning method that improves a model accuracy and generalization by randomly selecting subsets of the dataset and feature set for training. Prophet can automatically process various complex time series patterns and predict future trends and changes. LSTM can better capture long-term dependencies, avoid the vanishing or exploding gradients, and better handle long-term memory in sequence data. Fig. 6 shows the accuracy and recall of various models.

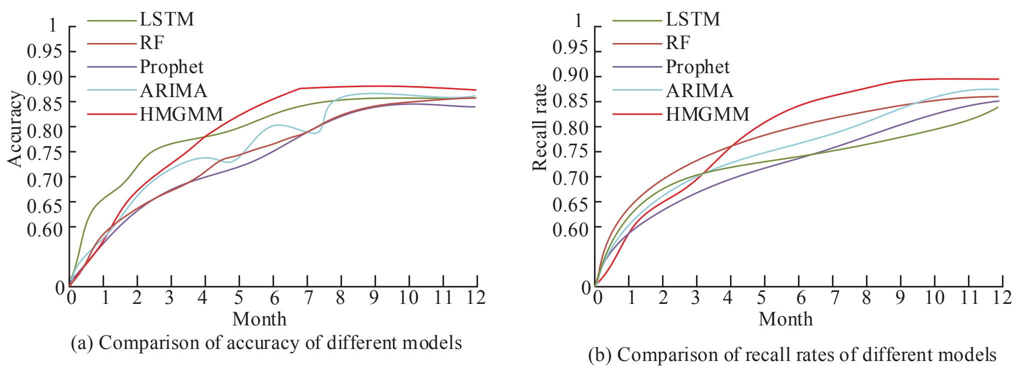


Figure 6: Accuracy and recall rate results of multiple models under time series

Fig. 6 (a) presents the RMSE of various models. The RMSE of LSTM and Prophet were both high, with values of 0.251 and 0.234, respectively. The RMSE of the proposed HMGMM was lower than 0.25 of ARIMA, indicating that the prediction error of HMGMM was smaller. Fig. 6 (b) compares recall rates for various models. From the results, the LSTM model had the worst recall rate, while HMGMM had the highest recall rate, reaching 0.901. This means that HMGMM is particularly good at capturing real movements and key turning points in the market. The higher recall rate makes HMGMM more sensitive to identify potential changes and abnormal events in the market, able to respond to major market fluctuations in a timely manner, and provide more accurate forecasting results.

Fig. 7 presents RMSE and MAE values.

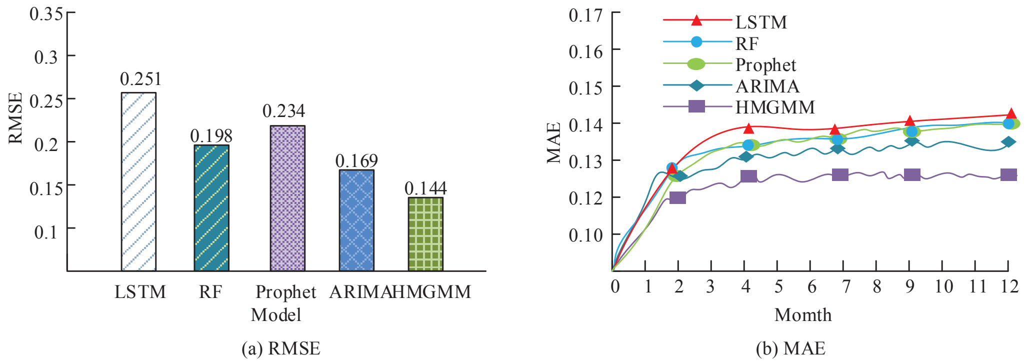


Figure 7: Comparison results of RMSE and MAE of various models

Fig. 7 (a) presents the RMSE of various models. The RMSE of LSTM and Prophet were both high, with values of 0.251 and 0.234, respectively. The RMSE of the proposed HMGMM was lower than 0.25 of ARIMA, indicating that the prediction error of HMGMM was smaller. Fig. 7 (b) shows the MAE results of multiple models. LSTM, Prophet, and RF had significant MAE values, indicating that these three models had significant prediction errors. The HMGMM error remained around 0.122, indicating that this prediction error was small and its prediction effect was good. From the perspective of method application, low RMSE and MAE are crucial for market forecasting models, especially in the rapidly changing and highly uncertain environment such as the AVs market. Smaller errors can provide more accurate market prediction for decision makers and help them make more accurate strategic decisions. Fig. 8 shows the MAPE of multiple models.

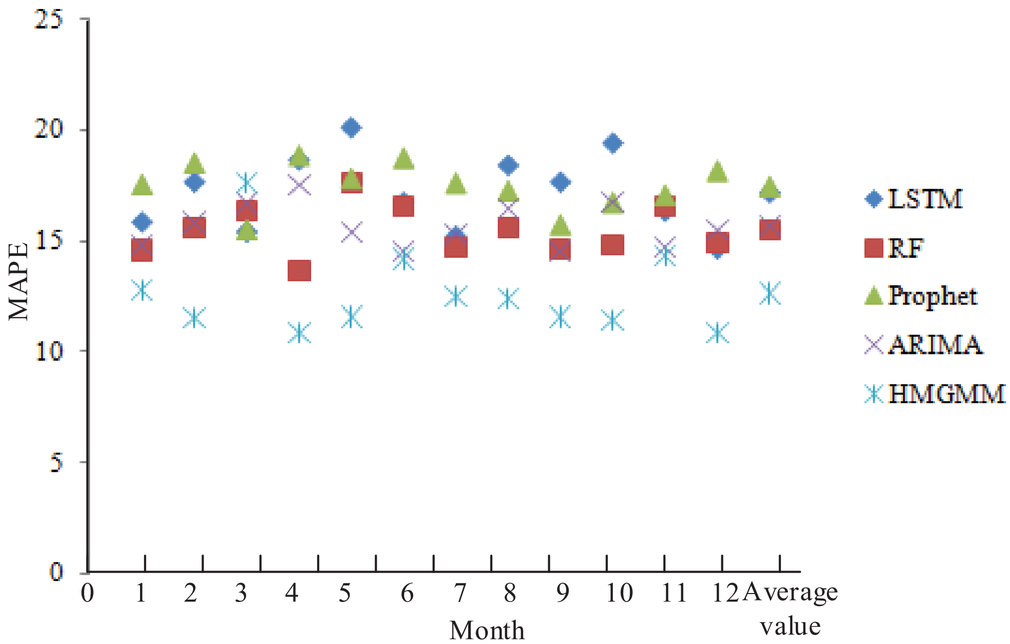


Figure 8: MAPE results for multiple models

Fig. 8 shows the MAPE comparison for multiple models. From the results, the MAPE value of RF and Prophet was relatively close, maintaining around 15, while the MAPE value of LSTM and RF was high, about 18. This indicates that there are significant errors in the prediction process, especially when there are significant changes in market dynamics, and their predictive ability is relatively limited. In contrast, HMGMM had the lowest MAPE value, which was basically maintained at about 12, and its average MAPE value was 12.77, which was the smallest among all models, indicating that HMGMM had higher accuracy and smaller error in actual prediction. Through comparison, it can be found that Prophet has the highest MAPE value and significant error, indicating that the model has a significant deviation in the fitting process of market data, resulting in a large gap between the predicted results and the actual values, and the fitting effect is poor. In contrast, HMGMM shows a more stable prediction effect, and its lower MAPE value reflects the better adaptability to market fluctuations and accurate prediction ability.

4.3 Discussion

The proposed application of HMGMM in AVs market forecasting shows that it has significant advantages in processing high-dimensional continuous data and dynamic market fluctuations. The innovation of HMGMM is that by introducing Gaussian mixture distribution, it can more accurately fit complex market dynamics, especially in multi-peak distributions and nonlinear fluctuations, showing stronger adaptability and interpretation ability. The experimental results showed that HMGMM was significantly better than HMM and other comparison models in accuracy (0.892), recall rate (0.901), F1 score (0.897), and other evaluation indicators. For example, compared with HMM, HMGMM not only has a higher accuracy rate, but also can better identify important changes and trends in the market. Therefore, faced with rapid market fluctuations, it is possible to respond and make predictions in a timely manner. This feature makes the HMGMM more applicable and reliable in AVs market prediction, especially under complex market conditions. From a practical application perspective, HMGMM can provide effective predictive support in highly dynamic and uncertain market environments. Especially for fast developing industries such as AVs, accurate market trend prediction can provide valuable information for decision-makers and help them adjust strategies timely.

5 Conclusion

This study uses the advantage of Probabilistic Big Data Analysis (BDA) to apply HMGMM to market forecasting in the field of AVs for the first time. The proposed model addresses the limitations inherent in traditional HMM by integrating Gaussian distributions, handling high latitude and continuous data and improving predictive accuracy.

The proposed HMGMM has undergone rigorous validation, demonstrating its effectiveness and superiority over HMM and other predictive models through comparative experiments. The model exhibits a higher overall fitting effect and stronger explanatory power, particularly at extreme points. The accuracy and recall rates reached 0.892 and 0.901, respectively, with an F1 score of 0.897. The proposed model outperforms other models, indicating its strong predictive ability and high accuracy in predicting market trends. In terms of performance metrics, the proposed model achieved lower values for RMSE, MAE, and MAPE, at 0.144, 0.122, and 12.77, respectively. These results underscore the model ability to accurately predict the AVs market, meeting the demands of market forecasting with higher precision compared to other models.

In conclusion, HMGMM is a reliable method for detecting hidden states in market predictions, which is recommended for application in similar prediction tasks. The proposed model can meet the needs of AVs market prediction and accurately predict the market. Although HMGMM demonstrates excellent predictive performance in this study, there are still some limitations that need further improvement. For example, this study only uses monthly data for model validation, and future research can attempt to apply the model to data with longer time spans to further examine its performance at different time scales. In addition, the sensitivity and robustness of HMGMM to abnormal data still need further research to improve its stability under extreme market conditions.

References

[1] K. A. Kowarski and H. Moors-Murphy. A review of big data analysis methods for baleen whale passive acoustic monitoring. *Marine Mammal Science*, 37(2):652–673, 2021. doi: 10.1111/mms.12758.

[2] R. J. Meredith, I. Carmichael, R. J. Woods, and A. S. Serianni. MA'AT analysis: Probability distributions of molecular torsion angles in solution from NMR spectroscopy. *Accounts of Chemical Research*, 56(17), Aug. 2023. doi: 10.1021/acs.accounts.3c00286.

[3] N. Chinthamu and M. Karukuri. Data science and applications. *Journal of Data Science and Intelligent Systems*, 1(2), Art. no. 2, Jul. 2023. doi: 10.47852/bonviewJDSIS3202837.

[4] O. Kulkarni, S. Jena, and V. Ravi Sankar. MapReduce framework based big data clustering using fractional integrated sparse fuzzy C means algorithm. *IET Image Processing*, 14(12):2719–2727, 2020. doi: 10.1049/iet-ipr.2019.0899.

[5] Y. Luo, S. Liu, L. Che, and Y. Yu. Analysis of temporal spatial distribution characteristics of PM2.5 pollution and the influential meteorological factors using big data in Harbin, China. *J Air Waste Manag Assoc*, 71(8):964–973, Aug. 2021. doi: 10.1080/10962247.2021.1902423.

[6] K. Garidis, L. Ulbricht, A. Rossmann, and M. Schmäh. Toward a user acceptance model of autonomous driving. In *Hawaii International Conference on System Sciences*, 2020. Available: https://api.semanticscholar.org/CorpusID:213678603.

[7] F. Hafeez et al. Autonomous vehicles perception, acceptance, and future prospects in the GCC: An analysis using the UTAUT-based model. *World Electric Vehicle Journal*, 15(5), Art. no. 5, May 2024. doi: 10.3390/wevj15050186.

[8] "Autonomous vehicle market size, share, trends | report [2030]." Accessed: Feb. 18, 2025. Available: https://www.fortunebusinessinsights.com/autonomous-vehicle-market-109045.

[9] A. Adewumi, S. E. Ewim, N. J. Sam-Bulya, and O. B. Ajani. Strategic innovation in business models: Leveraging emerging technologies to gain a competitive advantage. *International Journal of Management & Entrepreneurship Research*, 6(10), Art. no. 10, Oct. 2024. doi: 10.51594/ijmer.v6i10.1639.

[10] P. Grover, R. Debnath, D. Chaudhary, L. Bansal, and N. Batra. Role of artificial intelligence in autonomous vehicle. *SSRN Electronic Journal*, 2021. Available: https://api.semanticscholar.org/CorpusID:239622787.

[11] A. Kosovac, E. Muharemović, A. Čolaković, M. Lakača, and E. Šimić. Bosnia and herzegovina market research on the use of autonomous vehicles and drones in postal traffic. *Science, Engineering and Technology*, 1(2):32–37, Oct. 2021. doi: 10.54327/set2021/v1.i2.9.

[12] S. Liu and W. (David) Fan. Investigating the operational performance of connected and autonomous vehicles on signalized superstreets. *Transportation Planning and Technology*, Aug. 2021. Accessed: Feb. 17, 2025. Available: https://www.tandfonline.com/doi/full/10.1080/03081060.2021.1943130.

[13] S. Shin, Y. Cho, S. Lee, and J. Park. Assessing traffic-flow safety at various levels of autonomous-vehicle market penetration. *Applied Sciences*, 14(13):13, Jan. 2024. doi: 10.3390/app14135453.

[14] O. Panchenko, O. Balazyuk, T. Portovaras, V. Andrieieva, and V. Kotkovskyy. Analysis of financial statements as a business management tool. *AD ALTA: Journal of Interdisciplinary Research*, 14(1):157–161, Jan. 2024. doi: 10.33543/140139157161.

[15] C. Liu et al. Policy efforts to promote the adoption of autonomous vehicles: Subsidy and AVs lanes. *Journal of Advanced Transportation*, 2023(1):5658495, 2023. doi: 10.1155/2023/5658495.

[16] X.-W. Li and H.-Z. Miao. How to incorporate autonomous vehicles into the carbon neutrality framework of China: Legal and policy perspectives. *Sustainability*, 15(7), Jan. 2023. doi: 10.3390/su15075671.

[17] S. B. Salim and M. T. Husain. The role of artificial intelligence in autonomous vehicles: Challenges and opportunities. *International Journal For Multidisciplinary Research*, 2024. Available: https://api.semanticscholar.org/CorpusID:273343971.

[18] A. Rezaei, M. Cao, Q. Liu, and J. De Vos. Synthesising the existing literature on the market acceptance of autonomous vehicles and the external underlying factors. *Journal of Advanced Transportation*, 2023(1):6065060, 2023. doi: 10.1155/2023/6065060.

[19] E. Kim, Y. Kim, and J. Park. The necessity of introducing autonomous trucks in logistics 4.0. *Sustainability*, 2022.

[20] P. Stasinopoulos, N. Shiwakoti, and M. Beining. Use-stage life cycle greenhouse gas emissions of the transition to an autonomous vehicle fleet: A system dynamics approach. *Journal of Cleaner Production*, 2021.

[21] T. Mordan et al. Detecting 32 pedestrian attributes for autonomous vehicles. *arXiv preprint arXiv:2003.04567*, 2020.

[22] A. Bhardwaj, Y. Kumar, and N. Hasteer. A hybrid model to investigate perception towards adoption of autonomous vehicles in India. In *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, pages 1–5, 2021. doi: 10.1109/GUCON50781.2021.9573588.

[23] S. M. Pang, J. S. Ho, B. C. Tan, T. C. Lau, and N. Khan. Navigating the road to acceptance: Unveiling psychological and socio-demographic influences on autonomous vehicle adoption in malaysia. *Sustainability*, 16(18):18, Jan. 2024. doi: 10.3390/su16188262.

[24] H. Naderi and H. Nassiri. Examining the effect of moderating variables on autonomous public van acceptance model (APVAM). *PloS one*, 2023.

[25] K. Zheng, W. Xu, and X. Zhang. Multivariate regime switching model estimation and asset allocation. *COMPUTATIONAL ECONOMICS*, 2021. Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.1007_s10614-021-10203-9.

[26] J. Rapp, J. Tachella, Y. Altmann, S. McLaughlin, and V. K. Goyal. Advances in single-photon lidar for autonomous vehicles: Working principles, challenges, and recent advances. *IEEE Signal Processing Magazine*, 37(4):62–71, Jul. 2020. doi: 10.1109/MSP.2020.2983772.

[27] A. Boretti. A perspective on single-photon LiDAR systems. *Microwave and Optical Technology Letters*, 66(1):e33918, 2024. doi: 10.1002/mop.33918.

[28] T. Xu, M. W. Levin, and M. Cieniawski. A zone-based dynamic queueing model and maximum-stability dispatch policy for shared autonomous vehicles. *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021.

[29] M. N. Hasan Shuvo, Q. Zhu, and M. Hossain. Empowering digital twin: Early action decision through GAN-enhanced predictive frame synthesis for autonomous vehicles. In *2023 IEEE/ACM Symposium on Edge Computing (SEC)*, 2023.

[30] R. Roelofs et al. CausalAgents: A robustness benchmark for motion forecasting. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

[31] H. Jo, S.-A. Kim, and H. Kim. Forecasting the reduction in urban air pollution by expansion of market shares of eco-friendly vehicles: A focus on seoul, korea. *International Journal of Environmental Research and Public Health*, 2022.

[32] A. Arora et al. Forecasting based analysis of EV market. In *2022 22nd National Power Systems Conference (NPSC)*, 2022.

[33] J. H. Kim et al. Determinants of personal concern about autonomous vehicles. *Cities*, 2021.

[34] Y. Shi et al. Grid-centric traffic scenario perception for autonomous driving: A comprehensive review. *arXiv preprint arXiv:2304.05678*, 2023.

[35] S. Khandelwal et al. What-if motion prediction for autonomous driving. *arXiv preprint arXiv:2004.00657*, 2020.

[36] T. Feng, W. Wang, and Y. Yang. A survey of world models for autonomous driving. *arXiv preprint arXiv:2501.01234*, 2025.

[37] X. Wang, X. Lin, and M. Li. Aggregate modeling and equilibrium analysis of the crowdsourcing market for autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 132, p. 103362, Nov. 2021. doi: 10.1016/j.trc.2021.103362.

[38] A. Alatawneh and A. Torok. Potential autonomous vehicle ownership growth in hungary using the gompertz model. *Production Engineering Archives*, 29(2):155–161, Apr. 2023. doi: 10.30657/pea.2023.29.18.

[39] S. Fu and H. Fu. A method to predict electric vehicles' market penetration as well as its impact on energy saving and CO2 mitigation. *Science Progress*, 2021.

[40] Dan Chang, Rui Fan, and Yan Wang. Forecast of large earthquake emergency supplies demand based on PSO-BP neural network. *Teh. vjesn.*, 29(2), Apr. 2022, doi: 10.17559/TV-20211120092137.

[41] Zijian CHEN and Ji ZHAO. Bearing fault diagnosis based on wide deep convolutional neural network and long short term memory. *Teh. vjesn.*, 30(1), Feb. 2023, doi: 10.17559/TV-20220512135354.

[42] X. Ma, M. Li, J. Tong, and X. Feng. Deep learning combinatorial models for intelligent supply chain demand forecasting. *Biomimetics (Basel)*, 8(3):312, Jul. 2023, doi: 10.3390/biomimetics8030312.

[43] G. Avinash et al. Hidden markov guided deep learning models for forecasting highly volatile agricultural commodity prices. *Applied Soft Computing*, 158, p. 111557, Jun. 2024, doi: 10.1016/j.asoc.2024.111557.

[44] J. Anitha and N. Saranya. Cassava leaf disease identification and detection using deep learning approach. *INT J COMPUT COMMUN*, Int. J. Comput. Commun. Control, 17(2), Feb. 2022, doi: 10.15837/ijccc.2022.2.4356.

[45] S. Ettinger et al. Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[46] S. Safavi et al. Multi-sensor fault detection, identification, isolation and health forecasting for autonomous vehicles. *Sensors*, 21, 2021.

[47] H. Tabani, A. Balasubramaniam, E. Arani, and B. Zonooz. Challenges and obstacles towards deploying deep learning models on mobile devices. May 06, 2021, arXiv: arXiv:2105.02613. doi: 10.48550/arXiv.2105.02613.

[48] X. Liu et al. LiDAR-based 4D occupancy completion and forecasting. *arXiv preprint arXiv:2303.01234*, 2023.

[49] M. O. Berild and G.-A. Fuglstad. Non-stationary spatio-temporal modeling using the stochastic advection-diffusion equation. *arXiv preprint arXiv:2402.03456*, 2024.

[50] C. Ulrich et al. Mitigating false predictions in unreasonable body regions. Apr. 24, 2024, arXiv: arXiv:2404.15718. doi: 10.48550/arXiv.2404.15718.

[51] A. Daoudi and S. Mahmoudi. Enhancing brain segmentation in MRI through integration of hidden markov random field model and whale optimization algorithm. *Computers*, 13(5), p. 124, May 2024, doi: 10.3390/computers13050124.

[52] M. Zandawala, M. B. Amir, J. Shin, W. C. Yim, and L. A. Y. Guerra. Proteome-wide neuropeptide identification using NeuroPeptide-HMMer (NP-HMMer). Jul. 23, 2024, bioRxiv. doi: 10.1101/2024.07.20.604414.

[53] A. Xu. Theory based on HMM and bayesian analysis for predicting the remaining useful life of tools. In *Fourth International Conference on Mechanical, Electronics, and Electrical and Automation Control (METMS 2024)*, edited by Z. H. Khan, J. Zhang, and P. Zeng, SPIE, 2024, p. 131630B. doi: 10.1117/12.3030226.

[54] J. P. Williams, G. H. Hermansen, H. Strand, G. Clayton, and H. M. Nygård. Bayesian hidden markov models for latent variable labeling assignments in conflict research: Application to the role ceasefires play in conflict dynamics. *ARXIV-STAT.AP*, 2021, Available: https://www.paperdigest.org/paper/?paper_id=arxiv-2110.05475.

[55] M. Soleimani, F. Campean, and D. Neagu. Integration of hidden markov modelling and bayesian network for fault detection and prediction of complex engineered systems. *RELIAB. ENG. SYST. SAF.*, 2021, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.1016_j.ress.2021.107808.

[56] M. Diez, L. Burget, F. Landini, S. Wang, and H. Cernock. Optimizing bayesian hmm based X-vector clustering for the second dihard speech diarization challenge. *ICASSP*, 2020, Available: https://www.paperdigest.org/paper/?paper_id=icassp-ICASSP40776.2020.9053982-2020-04-26.

[57] W. Yang and L. Chen. Machine condition recognition via hidden semi-markov model. *COMPUT. IND. ENG.*, 2021, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.1016_j.cie.2021.107430.

[58] M. Lin, S. Wanqing, D. Chen, and E. Zio. Evolving connectionist system and hidden semi-markov model for learning-based tool wear monitoring and remaining useful life prediction. *IEEE ACCESS*, 2022, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.1109_access.2022.3196016.

[59] F. Jazayeri, A. Shahidinejad, and M. Ghobaei-Arani. A latency-aware and energy-efficient computation offloading in mobile fog computing: A hidden markov model-based approach. *J. SUPERCOMPUT.*, 2021, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.1007_s11227-020-03476-8.

[60] X. An. Prediction of stock price by hidden markov model. *BCP BUSINESS & MANAGEMENT*, 2023, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.54691_bcpbm.v38i.3709.

[61] B. Franzolini, A. Beskos, M. D. Iorio, W. P. Koziell, and K. Grzeszkiewicz. Change point detection in dynamic gaussian graphical models: The impact of COVID-19 pandemic on the US stock market. *ARXIV-STAT.ME*, 2022, Available: https://www.paperdigest.org/paper/?paper_id=arxiv-2208.00952.

[62] M.-Y. Li. Prediction of bitcoin price based on the hidden markov model. *PROCEEDINGS OF THE 2021 3RD INTERNATIONAL CONFERENCE ON ...*, 2021, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.2991_assehr.k.211209.481.

[63] S. Blazsek, A. Escribano, and A. Licht. Multivariate markov-switching score-driven models: An application to the global crude oil market. *STUDIES IN NONLINEAR DYNAMICS & ECONOMETRICS*, 2021, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.1515_snde-2020-0099.

[64] S. K. Kadhem and H. Thajel. Modelling of crude oil price data using hidden markov model. *THE JOURNAL OF RISK FINANCE*, 2023, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.1108_jrf-07-2022-0184.

[65] A. A. A. Danaa, M. I. Daabo, and A. Abdul-Barik. Detecting electronic banking fraud on highly imbalanced data using hidden markov models. *EARTHLINE JOURNAL OF MATHEMATICAL SCIENCES*, 2021, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.34198_ejms.7221.315332.

[66] H. Kui, J. Pan, R. Zong, H. Yang, W. Su, and W. Wang. [segmentation of heart sound signals based on duration hidden markov model]. *SHENG WU YI XUE GONG CHENG XUE ZA ZHI = JOURNAL OF ...*, 2020, Available: https://www.paperdigest.org/paper/?paper_id=pubmed-33140599.

[67] J. Chiu and A. Rush. Scaling hidden markov language models. *EMNLP*, 2020, Available: https://www.paperdigest.org/paper/?paper_id=emnlp-2020.emnlp-main.103-2020-11-12.

[68] A. J. Holbrook et al. Massive parallelization boosts big bayesian multidimensional scaling. *JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS: A JOINT ...*, 2020, Available: https://www.paperdigest.org/paper/?paper_id=pubmed-34168419.

[69] A. J. Holbrook et al. Massive parallelization boosts big bayesian multidimensional scaling. *J Comput Graph Stat*, 30(1):11–24, 2021, doi: 10.1080/10618600.2020.1754226.

[70] X. Ma and Y. Zhang. Digital innovation risk management model of discrete manufacturing enterprise based on big data analysis. *JOURNAL OF GLOBAL INFORMATION MANAGEMENT*, 2021, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.4018_jgim.286761.

[71] R. Chen, S. Yuan, C. Ma, H. Zhao, and Z.-Y. Feng. Tailored hidden markov model: A tailored hidden markov model optimized for cellular-based map matching. *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, 2022, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.1109_tie.2021.3135645.

[72] S. Murnani, N. A. Setiawan, and S. Wibirama. Robust object selection in spontaneous gaze-controlled application using exponential moving average and hidden markov model. *IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS*, 2024, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.1109_thms.2024.3413781.

[73] I. Sassi, S. Anter, and A. Bekkhoucha. A graph-based big data optimization approach using hidden markov model and constraint satisfaction problem. *JOURNAL OF BIG DATA*, 2021, Available: https://www.paperdigest.org/paper/?paper_id=doi.org_10.1186_s40537-021-00485-z.

[74] Z. Sheng et al. Graph-based spatial-temporal convolutional network for vehicle trajectory prediction in autonomous driving. *arXiv preprint arXiv:2102.03456*, 2021.

[75] K. Li. SWOT analysis of e-commerce development of rural tourism farmers' professional cooperatives in the era of big data. *IET Communications*, 16(5):592–603, 2022, doi: 10.1049/cmu2.12358.

[76] T. Lengauer. Statistical data analysis in the era of big data. *Chemie Ingenieur Technik*, 92(7):831–841, Jul. 2020, doi: 10.1002/cite.202000024.

[77] H. Wu et al. Scaling particle collision data analysis. *ARXIV-CS.LG*, 2024, Available: https://www.paperdigest.org/paper/?paper_id=arxiv-2412.00129.

[78] F. Wang et al. Smart households' aggregated capacity forecasting for load aggregators under incentive-based demand response programs. *IEEE Transactions on Industry Applications*, 56(2):1086–1097, Mar. 2020, doi: 10.1109/TIA.2020.2966426.

[79] D. W. Bunn, J. N. Inekwe, and D. MacGeehan. Analysis of the fundamental predictability of prices in the british balancing market. *IEEE Transactions on Power Systems*, 36(2):1309–1316, Mar. 2021, doi: 10.1109/TPWRS.2020.3015871.

[80] Thi Thanh Thao Tran and Q. Ma. Technology-enhanced self-regulation training: A dynamic training model to facilitate second language vietnamese learners' self-regulated writing skills. *System*, p. 103625, Feb. 2025, doi: 10.1016/j.system.2025.103625.

[81] A. Ngo and J. Kocoń. Integrating personalized and contextual information in fine-grained emotion recognition in text: A multi-source fusion approach with explainability. *Information Fusion*, 118, p. 102966, Jun. 2025, doi: 10.1016/j.inffus.2025.102966.

[82] P. Wang, R. Liu, X. Tian, X. Zhang, L. Qiao, and Y. Wang. Obstacle avoidance for environmentally-driven USVs based on deep reinforcement learning in large-scale uncertain environments. *Ocean Engineering*, 270, p. 113670, Feb. 2023, doi: 10.1016/j.oceaneng.2023.113670.

[83] J.-W. Park, Y.-M. Ju, and H.-S. Kim. Enhancing cold storage efficiency: Continuous deep deterministic policy gradient approach to energy optimization utilizing strategic sensor input data. *Energy Conversion and Management: X*, p. 100901, Feb. 2025, doi: 10.1016/j.ecmx.2025.100901.

[84] H. Amini, K. Alanne, and R. Kosonen. Building simulation in adaptive training of machine learning models. *Automation in Construction*, 165, p. 105564, 2024, doi: 10.1016/j.autcon.2024.105564.

[85] Y. Deng and Q. Lyu. Establishment of evaluation and prediction system of comprehensive state based on big data technology in a commercial blast furnace. *ISIJ Int.*, 60(5):898–904, May 2020, doi: 10.2355/isijinternational.ISIJINT-2019-545.

Copyright ©2025 by the authors. Licensee Agora University, Oradea, Romania.
This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.
Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

C O P E
Member since 2012
JM08090

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Chen, Xuemei; Qiu, Quanfeng; Chen, Zuanxu (2025). Enhanced Market Forecasting for Autonomous Vehicles Using a Hidden Mixture Gaussian Markov Model, *International Journal of Computers Communications & Control*, 20(5), 6911, 2025.
https://doi.org/10.15837/ijccc.2025.5.6911